MaSA-UNet: Manhattan Self-Attention UNet with Weighted Composed Loss for Skin Lesion Segmentation

Abel A. Reyes-Angulo
Michigan Technological University
1400 Townsend Dr, Houghton, MI 49931
areyesan@mtu.edu

Sidike Paheding
Fairfield University
1073 N Benson Rd, Fairfield, CT 06824
spaheding@fairfield.edu

Abstract

Melanoma is one of the deadliest forms of skin cancer in the USA, with a survival rate of 23% for delayed diagnosis. However, early detection could extend the survival rate up to 99%. Due to the importance of early analysis of skin lesions, many efforts have been dedicated to implementing end-to-end automated artificial intelligence systems to detect the presence of melanoma. In this study, we introduce the MaSA-UNet, a U-Net-like architecture complemented by the Manhattan Self-Attention mechanism for biomedical image segmentation. Additionally, we propose a set of weighted compound loss (WCL) functions and self-supervision learning mechanisms to improve the segmentation baseline performance of the MaSA-UNet deep learning model for skin lesion segmentation tasks, particularly focusing on melanoma segmentation. This research utilizes several popular and publicly available skin lesion datasets: the ISIC 2016, 2017, 2018, and PH^2 *datasets. The results showed that our proposed MaSA-UNet outperforms the state-of-the-art deep learning architectures for skin lesions segmentation tasks in terms of the Dice coefficient and Jaccard score.*

1. Introduction

Melanoma is known as one of the most serious and deadliest forms of skin cancer in the United States [14]. According to Siegel et al., [25], the number of deaths associated with this type of skin cancer is projected to be counted as 7,990, only in the USA during 2023. Although melanoma represents only 1% of skin cancer cases detected in the United States, with a survival rate of 23% for delayed diagnosis, it is responsible for the majority of fatalities related to skin cancer. It is worth mentioning, that early detection could lead to a high survival rate. Following initial surgery, a significant number of individuals with "thin melanoma" have a 5-year relative survival rate of 99% in the USA [17]. Therefore, accurate and efficient methods are required to guarantee a precise diagnosis of skin lesions, thus ensuring a timely treatment [1].

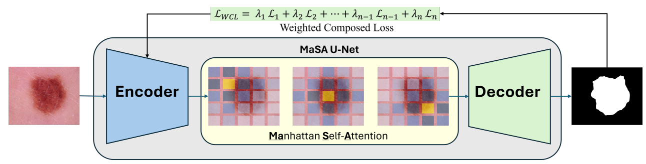

Figure 1. The proposed MaSA-UNet architecture for skin lesion segmentation. The Manhattan Self-Attention mechanism is used to highlight the most important features for the reconstruction of the segmentation mask

The use of deep learning (DL) to automate several tasks, ranging from software to industrial manufacturing, has been a trend in the last decade. Powered by large datasets, DL algorithms have proven to outperform human abilities in different tasks such as video games [19], board games [26], and object detection [24]. Biomedical image analysis for computer-assisted diagnosis is one of the benefited fields by DL [11], specifically due to the advancements in computer vision tasks such as image classification [12], object recognition [13], and image segmentation [4]. Image segmentation is well-known as a pixel-level classification task, which is usually adopted to precisely locate regions of interest, for instance, skin lesions. Performing skin lesion segmentation is a challenging task, in particular for cases where the lesions are not clearly contrasting with the skin color or have an almost imperceptible pigment [9]. Because of this and other scenarios related to biomedical imaging segmentation (BIS), several state-of-the-art DL architectures have been proposed [15], such as U-Net [23], V-Net [18], UPEN [21, 22], TransU-Net [4], and Swin-Unet [3].

In this work, we introduce a novel approach for extracting skin lesion regions using the Manhattan Self-Attention UNet (MaSA-UNet) combined with the Weighted Composed Loss (WCL) as an advanced DL architecture tailored for skin lesion segmentation tasks. This architecture is designed to enhance local context understanding through a two-dimensional decay attention mechanism, aiming to refine the accuracy of the predicted segmentation mask. Figure 1 highlights the central concept of MaSA-UNet. The goal of this study is to present MaSA-UNet not only as a novel architectural innovation but also as a means to produce precise segmentation masks for skin lesions. This is achieved by minimizing the difference between the predicted segmentation mask and the manually annotated ground truth. We introduce the Weighted Composed Loss (WCL) function with the anticipation that a combination of weighted loss functions will enhance MaSA-UNet's performance in segmenting Melanoma and accurately identify significant features within the skin lesions.

In summary, our contributions to this work are outlined as follows:

- To our knowledge, this is the inaugural study to introduce the Manhattan Self-Attention mechanism for medical image segmentation tasks, marking a novel application in the field.
- We propose MaSA-UNet, a novel end-to-end DL model for skin lesion segmentation, incorporating the Weighted Composed Loss that blends multiple weighted loss functions to enhance segmentation accuracy. MaSA-UNet achieves competing performance compared to existing models and will be made available to the research community through a repository link.

The rest of this paper is organized as follows: Section 2 provides a detailed discussion of the proposed methodology. In section 3, qualitative and quantitative results are analyzed. Finally, section 4 concludes this work, and provides suggestions for future directions.

2. Methodology

2.1. MaSA: The Manhattan Self-Attention

Introduced by Fan et al. [7] and inspired by the Retentive Network [27], the Manhattan Self-Attention (MaSA) mechanism proposes a novel approach within the scope of enhancing the operational efficiency and performance of attention-based models. This method diverges from conventional attention mechanisms by adopting a unique configuration that significantly diminishes computational demands without compromising accuracy. It simplifies the process of attention computation, focusing on a more efficient utilization of computational resources. This advancement is particularly advantageous for handling extensive datasets or intricate pattern recognition tasks, making it a valuable asset for progressing research in areas such as natural language processing and computer vision. MaSA is formulated as follows

$$BiRetention(X) = (QK^\top \odot D^{Bi})V \tag{1}$$

where $D^{\mathrm{Bi}}_{nm} = \gamma^{|n-m|}$ represents the interaction between the n-th and m-th tokens.

For two-dimensional data, such as images, MaSA extends the decay matrix to incorporate Manhattan distances

$$D^{2d}_{nm} = \gamma^{|x_n - x_m| + |y_n - y_m|} \tag{2}$$

The core MaSA equation, accounting for spatial decay, is

$$MaSA(X) = (Softmax(QK^\top) \odot D_{2d})V \tag{3}$$

where the attention scores are calculated separately for horizontal and vertical directions

$$MaSA(X) = Attn_H(Attn_W V)^\top \tag{4}$$

with $Attn_H = Softmax(Q_H K_H^\top) \odot D_H$ and $Attn_W = Softmax(Q_W K_W^\top) \odot D_W$.

The Local Context Enhancement module is then introduced for further enhancing the local expression capability of MaSA, thus the final model is expressed as

$$X_{out} = MaSA(X) + LCE(V) \tag{5}$$

Here, LCE employs depthwise convolutions, underlining MaSA's utility in leveraging spatial relationships for enhanced attention modeling in vision tasks.

2.2. The Weighted Composed Loss (WCL)

The Weighted Composed Loss (WCL) function is an optimization approach introduced alongside the MaSA-UNet architecture, aimed at improving the segmentation of biomedical images. It leverages a combination of distinct loss functions, each contributing unique benefits to the segmentation task. The essence of WCL lies in its capability to systematically merge these functions using specific weights (λ), thereby tailoring the loss landscape to more effectively pinpoint the global minima.

A generalized expression of WCL incorporating n loss functions is given by:

$$WCL_n = \sum_{i=1}^{n} \lambda_i \mathcal{L}_i \tag{6}$$

where λ_i denotes the weight assigned to each loss function $\mathcal{L}_i$, optimized to achieve superior performance.

The WCL integrates several key loss functions, including:

Binary Cross Entropy Loss (BCE), fundamental for binary classification, defined as:

$$\mathcal{L}_{BCE} = -[y \log(\hat{y}) + (1-y)\log(1-\hat{y})] \tag{7}$$

Dice Coefficient Loss, crucial for image segmentation, given by:

$$\mathcal{L}_{Dice} = 1 - \frac{2 \cdot \sum_{i=1}^{N} y_i \cdot \hat{y}_i}{\sum_{i=1}^{N} y_i + \sum_{i=1}^{N} \hat{y}_i} \tag{8}$$

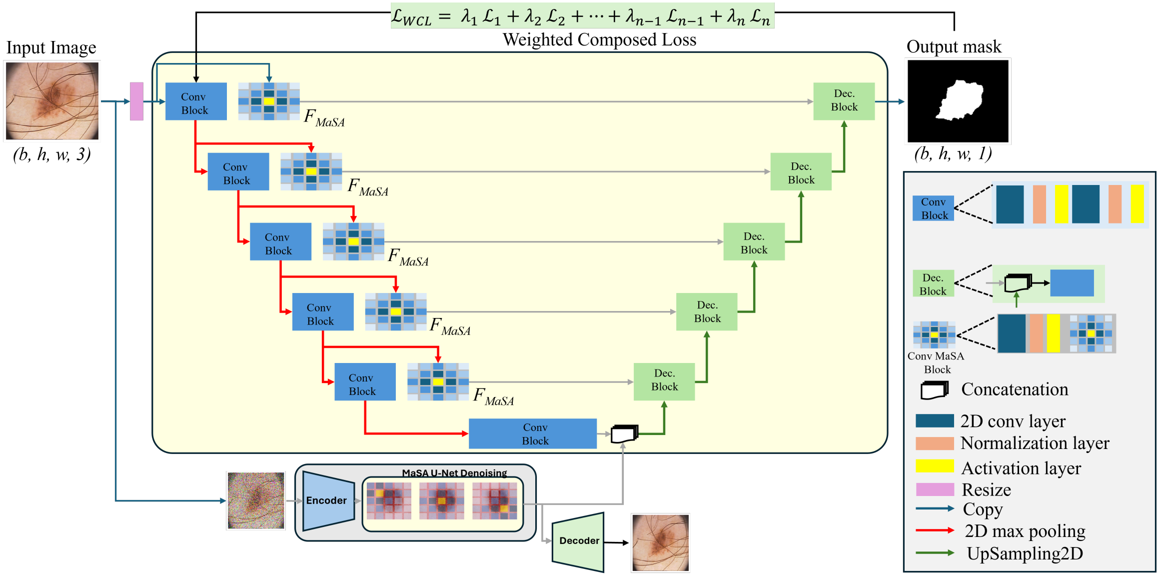

Figure 2. Detailed illustration of the proposed MaSA-UNet model. The picture depicts the key components of the architecture, the pre-trained MaSA denoised model, the MaSA segmentation model, and the WCL function.

Jaccard Loss (IoU Loss), measuring the overlap between predicted and ground truth masks:

$$\mathcal{L}_{Jaccard} = 1 - \frac{\sum_{i=1}^{N} y_i \cdot \hat{y}_i}{\sum_{i=1}^{N} y_i + \sum_{i=1}^{N} \hat{y}_i - \sum_{i=1}^{N} y_i \cdot \hat{y}_i} \tag{9}$$

Focal Loss, addressing class imbalance by focusing more on difficult, misclassified examples:

$$\mathcal{L}_{Focal} = -\alpha_t (1 - p_t)^{\gamma} log(p_t) \tag{10}$$

2.3. The MaSA-UNet: Manhattan Self-Attention UNet with Weighted Composed Loss

The MaSA-UNet architecture innovatively integrates the Manhattan Self-Attention (MaSA) mechanism, a self-supervised encoder pre-training strategy, and WCL to enhance the accuracy of skin lesion segmentation. This architecture capitalizes on the MaSA mechanism's ability to efficiently highlight relevant features within the image through spatial attention, significantly improving the model's focus on pertinent areas for accurate segmentation.

The self-supervised encoder pre-training further refines the model's ability to discern intricate patterns within biomedical images, allowing for a more nuanced understanding and representation of the data. This pre-training phase utilizes a denoising task to learn a robust feature representation that is highly beneficial for the segmentation task.

The WCL function is meticulously designed to combine multiple loss functions with distinct weights (λ), optimizing the model's performance by addressing the specific challenges posed by skin lesion segmentation. The weighted approach allows for a tailored optimization strategy that emphasizes the most relevant aspects of the segmentation task, ensuring high fidelity in the generated masks relative to the ground truth.

Mathematically, the MaSA-UNet can be represented as follows

$$MaSA - UNet(X) = F_{\text{MaSA}}(E_{\text{pre-trained}}(X)) \tag{11}$$

where X is the input image, F_{MaSA} denotes the function representing the application of the Manhattan Self-Attention mechanism in the framework of UNet architecture[23], and $E_{\text{pre-trained}}$ indicates the pre-trained encoder function.

The WCL for the model is defined as

$$\mathcal{L}_{WCL} = \sum_{i=1}^{n} \lambda_i \mathcal{L}_i(MaSA - UNet(X), Y) \tag{12}$$

where Y is the ground truth mask, $\mathcal{L}_i$ represents the individual loss functions (e.g., BCE, Dice, Jaccard, Focal), and λ_i are the corresponding weights determined through optimization to yield the best performance.

Through meticulous design and integration, the MaSA-UNet achieves competitive performance for skin lesion segmentation. Figure 2 shows details of the design of the proposed MaSA-UNet architecture.

3. Experimental Setup and Results

3.1. Datasets

In this study, the proposed method is trained and tested over four publicly available skin lesion datasets, including PH^2 [1]

[1] https://www.fc.up.pt/addi/ph2%20database.html

Table 1. Experimental results over the test split of the different benchmark datasets. '-': No results reported. key: [**Best**, **Second Best**]

Method	ISIC 2016		ISIC 2017		ISIC 2018		PH2	
	JS	DC	JS	DC	JS	DC	JS	DC
UNet [23]	0.812	0.887	0.658	0.794	0.685	0.813	0.794	0.873
UNet++ [30]	0.818	0.889	0.686	0.814	0.714	0.833	0.813	0.890
Attention Unet [20]	0.811	0.886	0.688	0.815	0.730	0.844	0.802	0.880
CA-Net [8]	0.807	0.881	-	-	-	-	0.752	0.847
SegFormer [29]	-	-	**0.825**	**0.904**	**0.842**	**0.914**	-	-
Swin Unet [2]	-	-	0.741	0.851	0.768	0.869	-	-
TransUNet[4]	**0.849**	**0.913**	0.775	0.873	0.825	0.904	**0.840**	**0.910**
MaSA-UNet (ours)	**0.860**	**0.925**	**0.830**	**0.907**	**0.836**	**0.911**	**0.879**	**0.936**

[16], ISIC 2016 [10], ISIC 2017 [6], and ISIC 2018 [5, 28][2] datasets.

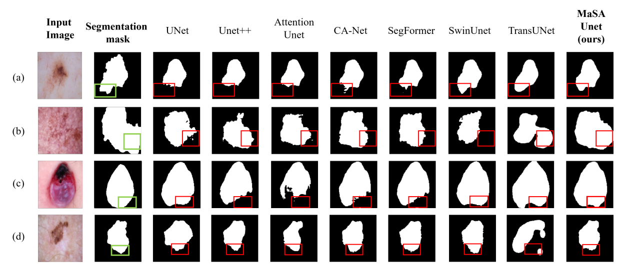

Figure 3. Qualitative results over the (a) ISIC 2016, (b) ISIC 2017, (c) ISIC 2018, and PH^2 datasets, comparing the MaSA-UNet with state-of-the-art image segmentation models. Green boxes indicate the region of interest in the GT. Red boxes indicate the comparison with the GT on each output.

3.2. Implementation details

The ISIC 2016 dataset includes 900 and 379 samples for training and testing respectively, each sample with the corresponding manually annotated segmentation mask. For the experiments we used 810 samples for training the DL architecture, 90 samples for validation and hyperparameter selection, including the λ values for the $\mathcal{L}_{WCL}$, and the remaining 379 samples for final testing evaluation. In the case of ISIC 2017 and 2018, each dataset contains a total of 2,750 and 3,964 samples respectively. The PH^2 dataset contains a total of 200 samples, randomly split in the proportion of 7:1:2 for training, validation, and testing respectively. To alleviate the computational resources constraint, the size of the input images and the corresponding segmentation mask were resized to a uniform resolution of 224×224 on all datasets. The proposed MaSA-UNet architecture preserves the same layout as the vanilla UNet architecture (Initial filter size of 32 and duplicating every downsample layer in the encoder, opposite behavior in the decoder size), with the incorporation of the key components as the MaSA block and the pre-trained denoise model. A modified version of the vanilla U-Net architecture was used to perform the grid search for the proposed $\mathcal{L}_{WCL}$ function. Experiments without using $\mathcal{L}_{WCL}$, used the $\mathcal{L}_{Jaccard}$ as default loss. The pre-trained MaSA-based denoising model runs for 250 epochs, while during the grid search, for the λ values, each experiment runs for a total of 50 epochs. On the other hand, the final training process of the MaSA-UNet is trained for a total of 150 epochs, monitoring the best performance over the validation set. The batch size while training is set to 16 for all the benchmark datasets. We use the Adam optimizer for all the experiments with an initial learning rate of 1×10^{-3} and a scheduled reducer factor of 0.01 on every epoch after the first 20 epochs.

[2]https://challenge.isic-archive.com/data/

3.3. Experimental Results

Quantitative results summarized in table 1 show that the MaSA-UNet DL segmentation model significantly improves skin lesion segmentation performance over competing models in terms of Dice coefficient (DC) and Jaccard score (JS). Our comprehensive study compares MaSA-UNet to baseline models (UNet, UNet++), attention-based models (Attention UNet, CA-Net), and transformer-based models (SegFormer, Swin UNet, TransUNet), highlighting MaSA-UNet's superior capabilities. MaSA-UNet outperforms state-of-the-art DL models in three benchmark datasets (ISIC 2016, ISIC 2017, and PH^2) and competes closely with SegFormer on the ISIC 2018 dataset, surpassing other models in performance. Qualitatively, a few samples from the different datasets are shown in Figure 3, contrasting the performance of the proposed MaSA UNet with the competing segmentation models.

4. Conclusion

In this work, we introduced a novel DL architecture, MaSA-UNet, aimed at enhancing the accuracy of skin lesion segmentation. The MaSA-UNet incorporates a Manhattan Self-Attention UNet model combined with a Weighted Composed Loss (WCL). This configuration integrates the strengths of the bidirectional Manhattan Self-Attention mechanism with different loss functions, assigned varying weights, to improve segmentation performance as reflected in the Jaccard score and Dice coefficient metrics. Our experimental results demonstrate that MaSA-UNet can accurately delineate skin lesion regions, reducing the discrepancy between the predicted segmentation mask and the ground truth, and enhancing overall segmentation accuracy.

In future research, we plan to enhance the efficiency of the weight-search process to decrease both time consumption and computational expense. We also aim to further investigate the applicability of the Manhattan Self-Attention mechanism across various computer vision tasks. Furthermore, efforts will be dedicated to advancing the interpretability of the segmentation outcomes facilitated by MaSA-UNet.

References

[1] Bhavneet Bhinder, Coryandar Gilvary, Neel S Madhukar, and Olivier Elemento. Artificial intelligence in cancer research and precision medicine. *Cancer discovery*, 11(4): 900–915, 2021. 1

[2] Hu Cao, Yueyue Wang, Joy Chen, Dongsheng Jiang, Xiaopeng Zhang, Qi Tian, and Manning Wang. Swin-unet: Unet-like pure transformer for medical image segmentation. In *European conference on computer vision*, pages 205–218. Springer, 2022. 4

[3] Hu Cao, Yueyue Wang, Joy Chen, Dongsheng Jiang, Xiaopeng Zhang, Qi Tian, and Manning Wang. Swin-unet: Unet-like pure transformer for medical image segmentation. In *Computer Vision–ECCV 2022 Workshops: Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part III*, pages 205–218. Springer, 2023. 1

[4] Jieneng Chen, Yongyi Lu, Qihang Yu, Xiangde Luo, Ehsan Adeli, Yan Wang, Le Lu, Alan L Yuille, and Yuyin Zhou. Transunet: Transformers make strong encoders for medical image segmentation. *arXiv preprint arXiv:2102.04306*, 2021. 1, 4

[5] Noel Codella, Veronica Rotemberg, Philipp Tschandl, M Emre Celebi, Stephen Dusza, David Gutman, Brian Helba, Aadi Kalloo, Konstantinos Liopyris, Michael Marchetti, et al. Skin lesion analysis toward melanoma detection 2018: A challenge hosted by the international skin imaging collaboration (isic). *arXiv preprint arXiv:1902.03368*, 2019. 4

[6] Noel CF Codella, David Gutman, M Emre Celebi, Brian Helba, Michael A Marchetti, Stephen W Dusza, Aadi Kalloo, Konstantinos Liopyris, Nabin Mishra, Harald Kittler, et al. Skin lesion analysis toward melanoma detection: A challenge at the 2017 international symposium on biomedical imaging (isbi), hosted by the international skin imaging collaboration (isic). In *2018 IEEE 15th international symposium on biomedical imaging (ISBI 2018)*, pages 168–172. IEEE, 2018. 4

[7] Qihang Fan, Huaibo Huang, Mingrui Chen, Hongmin Liu, and Ran He. Rmt: Retentive networks meet vision transformers. *arXiv preprint arXiv:2309.11523*, 2023. 2

[8] Ran Gu, Guotai Wang, Tao Song, Rui Huang, Michael Aertsen, Jan Deprest, Sébastien Ourselin, Tom Vercauteren, and Shaoting Zhang. Ca-net: Comprehensive attention convolutional neural networks for explainable medical image segmentation. *IEEE transactions on medical imaging*, 40(2): 699–711, 2020. 4

[9] Yonis Gulzar and Sumeer Ahmad Khan. Skin lesion segmentation based on vision transformers and convolutional neural networks—a comparative study. *Applied Sciences*, 12(12): 5990, 2022. 1

[10] David Gutman, Noel CF Codella, Emre Celebi, Brian Helba, Michael Marchetti, Nabin Mishra, and Allan Halpern. Skin lesion analysis toward melanoma detection: A challenge at the international symposium on biomedical imaging (isbi) 2016, hosted by the international skin imaging collaboration (isic). *arXiv preprint arXiv:1605.01397*, 2016. 4

[11] Mohammad H Jafari, Nader Karimi, Ebrahim Nasr-Esfahani, Shadrokh Samavi, S Mohamad R Soroushmehr, K Ward, and Kayvan Najarian. Skin lesion segmentation in clinical images using deep learning. In *2016 23rd International conference on pattern recognition (ICPR)*, pages 337–342. IEEE, 2016. 1

[12] Qing Li, Weidong Cai, Xiaogang Wang, Yun Zhou, David Dagan Feng, and Mei Chen. Medical image classification with convolutional neural network. In *2014 13th international conference on control automation robotics & vision (ICARCV)*, pages 844–848. IEEE, 2014. 1

[13] Zhuoling Li, Minghui Dong, Shiping Wen, Xiang Hu, Pan Zhou, and Zhigang Zeng. Clu-cnns: Object detection for medical images. *Neurocomputing*, 350:53–59, 2019. 1

[14] Miguel A Linares, Alan Zakaria, and Parminder Nizran. Skin cancer. *Primary care: Clinics in office practice*, 42(4):645–659, 2015. 1

[15] Priyanka Malhotra, Sheifali Gupta, Deepika Koundal, Atef Zaguia, and Wegayehu Enbeyle. Deep neural networks for medical image segmentation. *Journal of Healthcare Engineering*, 2022, 2022. 1

[16] Teresa Mendonça, Pedro M Ferreira, Jorge S Marques, André RS Marcal, and Jorge Rozeira. Ph 2-a dermoscopic image database for research and benchmarking. In *2013 35th annual international conference of the IEEE engineering in medicine and biology society (EMBC)*, pages 5437–5440. IEEE, 2013. 4

[17] Kimberly D Miller, Leticia Nogueira, Angela B Mariotto, Julia H Rowland, K Robin Yabroff, Catherine M Alfano, Ahmedin Jemal, Joan L Kramer, and Rebecca L Siegel. Cancer treatment and survivorship statistics, 2019. *CA: a cancer journal for clinicians*, 69(5):363–385, 2019. 1

[18] Fausto Milletari, Nassir Navab, and Seyed-Ahmad Ahmadi. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In *2016 fourth international conference on 3D vision (3DV)*, pages 565–571. Ieee, 2016. 1

[19] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Andrei A Rusu, Joel Veness, Marc G Bellemare, Alex Graves, Martin Riedmiller, Andreas K Fidjeland, Georg Ostrovski, et al. Human-level control through deep reinforcement learning. *nature*, 518(7540):529–533, 2015. 1

[20] Ozan Oktay, Jo Schlemper, Loic Le Folgoc, Matthew Lee, Mattias Heinrich, Kazunari Misawa, Kensaku Mori, Steven McDonagh, Nils Y Hammerla, Bernhard Kainz, et al. Attention u-net: Learning where to look for the pancreas. *arXiv preprint arXiv:1804.03999*, 2018. 4

[21] Sidike Paheding, Abel A Reyes, Mohammad Alam, and Vijayan K Asari. Medical image segmentation using u-net and progressive neuron expansion. In *Pattern Recognition and Tracking XXXIII*, page 1210102. SPIE, 2022. 1

[22] Sidike Paheding, Abel A Reyes, Jeevan S Devagiri, Vijayan K Asari, and Mohammad Alam. U-pen++: redesigning u-pen architecture with multi-head attention for retinal image segmentation. In *Pattern Recognition and Tracking XXXIV*, page 1252702. SPIE, 2023. 1

[23] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmen-

tation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*, pages 234–241. Springer, 2015. 1, 3, 4

[24] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, et al. Imagenet large scale visual recognition challenge. *International journal of computer vision*, 115:211–252, 2015. 1

[25] Rebecca L Siegel, Kimberly D Miller, Nikita Sandeep Wagle, and Ahmedin Jemal. Cancer statistics, 2023. *CA: a cancer journal for clinicians*, 73(1):17–48, 2023. 1

[26] David Silver, Aja Huang, Chris J Maddison, Arthur Guez, Laurent Sifre, George Van Den Driessche, Julian Schrittwieser, Ioannis Antonoglou, Veda Panneershelvam, Marc Lanctot, et al. Mastering the game of go with deep neural networks and tree search. *nature*, 529(7587):484–489, 2016. 1

[27] Yutao Sun, Li Dong, Shaohan Huang, Shuming Ma, Yuqing Xia, Jilong Xue, Jianyong Wang, and Furu Wei. Retentive network: A successor to transformer for large language models. *arXiv preprint arXiv:2307.08621*, 2023. 2

[28] Philipp Tschandl, Cliff Rosendahl, and Harald Kittler. The ham10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. *Scientific data*, 5(1):1–9, 2018. 4

[29] Enze Xie, Wenhai Wang, Zhiding Yu, Anima Anandkumar, Jose M Alvarez, and Ping Luo. Segformer: Simple and efficient design for semantic segmentation with transformers. *Advances in Neural Information Processing Systems*, 34:12077–12090, 2021. 4

[30] Z Zhou, MMR Siddiquee, N Tajbakhsh, and J UNet+ Liang. A nested u-net architecture for medical image segmentation (2018). *arXiv preprint arXiv:1807.10165*. 4